

Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Fundamentos de Big Data

Lección 4

Introducción a Data Mining

ÍNDICE

- | Introducción
- | Objetivos
- | Competición Kaggle: Titanic Dataset (para aprendizaje)
- | `pd.crosstab()`
- | `sns.violinplot()`

INTRODUCCIÓN

En esta lección haremos una breve introducción al Data Mining

Hay varios conceptos importantes de esta lección que trataremos de ir explicando.

OBJETIVOS

Al finalizar esta lección serás capaz de:

- 1 Crear nuevos DataFrames bajo ciertas premisas en algunas columnas
- 2 Hacer gráficos de los cuales obtener información de un Dataset
- 3 Conocer un poco mejor Kaggle, como plataforma para aprender Data Science

Competición Kaggle: Titanic Dataset (para aprendizaje)

Please read the competition rules



Titanic - Machine Learning from Disaster Rules

By clicking on the "I Understand and Accept" button below, you agree to be bound by the competition rules.

I Understand and Accept

Join Competition

pd.crosstab()

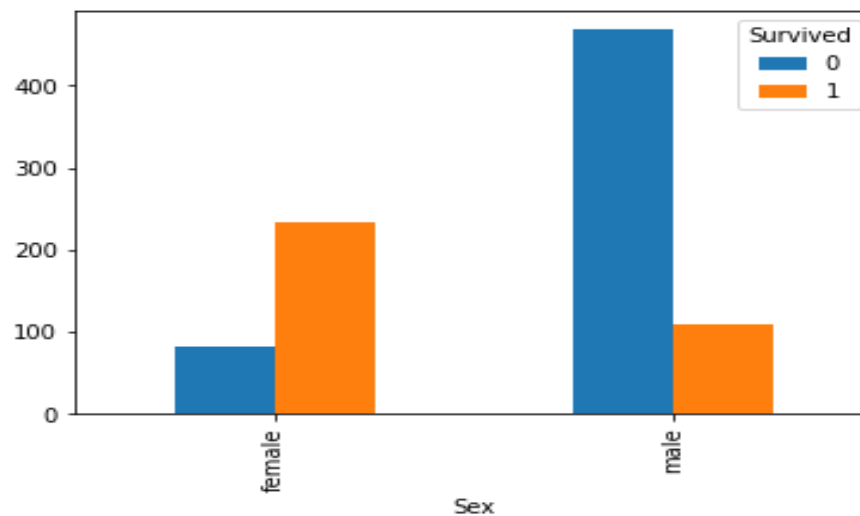
Forma 3 de seleccionar información concreta: crosstab

```
In [20]: pd.crosstab(df.Sex, df.Survived)
```

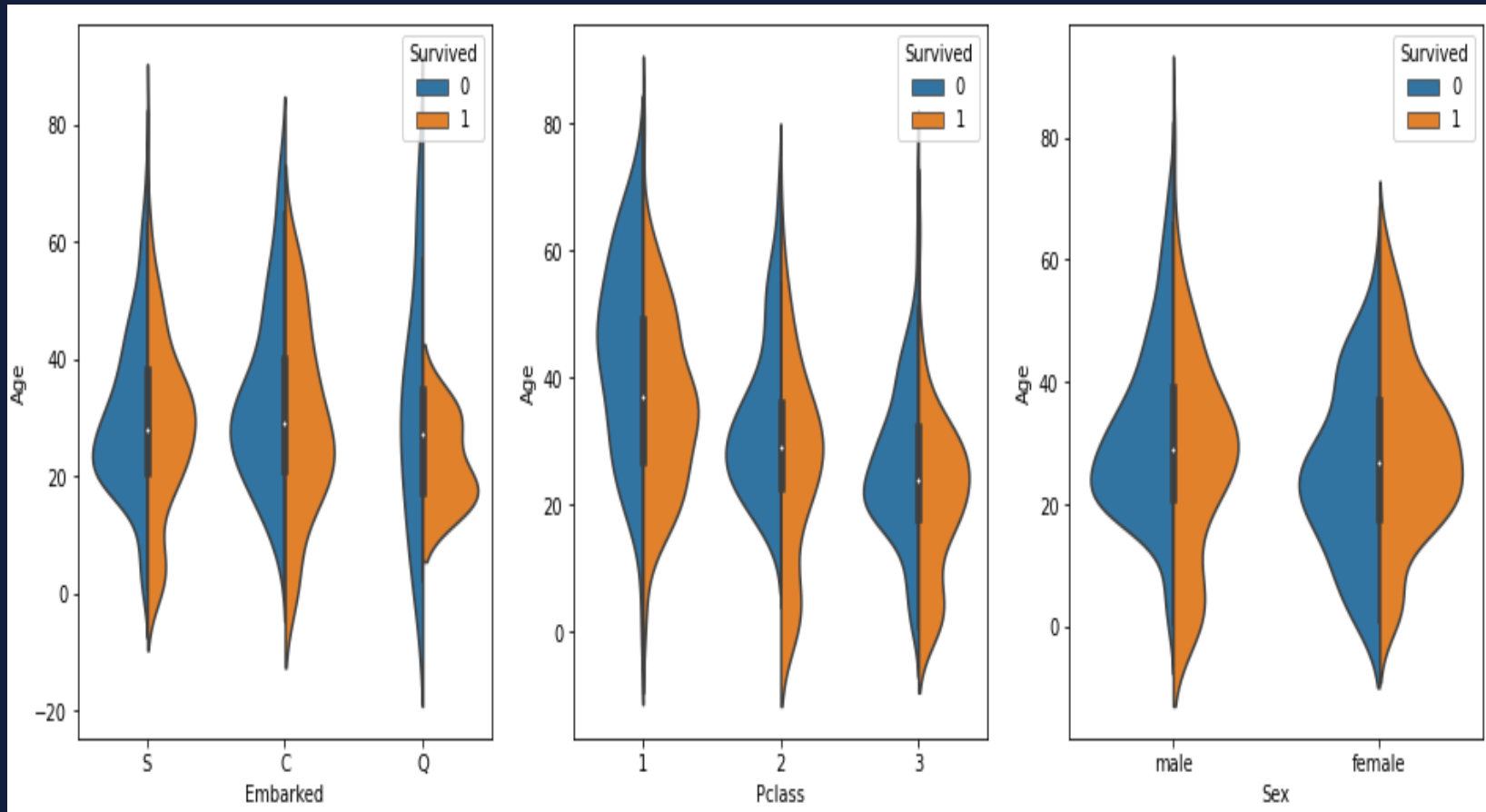
```
Out[20]:
```

Survived	0	1
Sex		
female	81	233
male	468	109

```
In [21]: pd.crosstab(df.Sex, df.Survived).plot(kind="bar")  
plt.show()
```



sns.violinplot()



MUCHAS GRACIAS POR SU ATENCIÓN



jmpena@grupomainjobs.com



José Manuel Peña

<https://www.linkedin.com/in/jos%C3%A9-manuel-pe%C3%B1a-castro-7566b349/>



twitter.com/eiposgrados



facebook.com/eiposgrados



instagram.com/eiposgrados